DOCUMENT RESUME

ED 110 481 TM 004 754

AUTHOR Hecht, Kathryn A.
TITLE Overview of Problems Involved in Validating Professional Licensing and Certification Examinations.
PUB DATE [Apr 74]
NOTE 22p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Chicago, Illinois, April 16-18, 1974)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS *Certification; *Nurses; Performance Tests; Predictive Validity; Professional Occupations; Selection; State Licensing Boards; Testing; Testing Problems; *Test Validity

ABSTRACT

A large amount of professional interest has been focused upon the ambiguities and problems involved in the conduct of professional licensing and certification through examinations. What seems to be a simple problem on the surface, that being the policing of professionals for competence and the practice of conducting this policing so that it offers equal fairness to all, turns out to be a very complex problem involving unresolved conceptual, legal, and methodological issues particularly with examination validity. There are four main areas of concern: (1) criticism of testing, (2) the growing number of jobs requiring licensing, (3) discriminatory practices in hiring and occupations access, and (4) validity of certification through testing. The example used is the field of nursing. (Author/DEP)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED110481

OVERVIEW OF PROBLEMS INVOLVED IN VALIDATING PROFESSIONAL LICENSING AND CERTIFICATION EXAMINATIONS

Kathryn A. Hecht, Ed.D.
Center for Northern Educational Research
University of Alaska
Fairbanks, Alaska 99701

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Presented at the 1974 Annual Meeting of the
National Council on Measurement in Education
Symposium (Session 7A)
Validation of Professional Licensing and Certification Examinations:
A Methodological Dilemma
Chicago, 1974

TM 004 754

How does an evaluator from Alaska come to be addressing you today on validation of licensing exams? Last year, while working as an independent consultant, I was asked by the National League for Nursing to do a background paper on the validation of the RN (Registered Nurse) licensing examination and related work on performance testing. Naively, I thought it would be a simple task of pulling together what had been done in other professions. It turned out to be a much more complex and interesting task than I had expected, and questions and concerns raised during that study led directly to our meeting together today. (One of our participants, Paul Jacobs, is now validation study director for the NLN and he will tell you more about the specifics of that effort.)

Definitions of Licensure and Certification

First, as part of an overview there is a simple matter of defining licensure and certification, . . . only it is not so simple. There is no standard definition nor usage of the terms. For the purposes of our discussion I think the most useful definitions are those proposed by O. Jensen (1972). In an unpublished paper, he discusses licensure and certification as two types of minimum competency testing, in that the purpose of the tests is to establish an individual's status with respect to an established go/no-go criteria. Licensing is usually a mandatory program designed to protect the public from incompetent practitioners, that is, to prevent an individual with particular deficiencies from entering practice. Jensen call this "selecting out". Certification, on the other hand, is usually a voluntary program with the emphasis on granting special status to an individual with more than run-of-the-mill knowledge, ability, and skill. This Jensen calls "selecting-in".

Perhaps the best known example of a "selecting-out" exam would be a driving license, where the public is protected from those whose driving knowledge is judged not to be up to standard. Another example is the RN exam, whose espoused purpose is to "measure minimum safety and effectiveness of practice, for the protection of the public" (N.L.N. 1961). Both licenses represent a legal right to engage in the appropriate activity.

Examples of "selecting-in" or certification are the "diplomate" program for medical specialities and the new certification program for automobile mechanics. They are both exams designed for experienced practitioners which provide evidence of superior capability in a specialty within the occupation.

Since validation deals with the purpose to which the test is intended, I believe these interpretations and distinctions to be important for our discussion. It should be obvious that the same test could not serve both licensure and certification purposes as defined here.

Unfortunately, neither this distinction nor any other I can locate fits current usages of the terms. For example, teacher certification I believe to be a misnomer. It is a legal requirement to begin teaching, to protect the public from incompetence and signifies no special standing within the profession. I am sure you can think of other cases which do not fit the given definition.

Why the Concern?

Next, a brief look at why the growing concern about licensing at this time? There are four concerns I will outline briefly. (Several of the participants and discussants are especially well qualified to discuss them further.)

The first is the criticism of testing in general which in the past decade has become a popular cause making frequent headlines and even best sellers (Hoffmann, B., 1962)

Second, there has been a proliferation of jobs requiring licensing and a hodgepodge of local and state legislated bodies emerging to control the process. Benjamin Shimberg (one of our discussants) and others (1972) have written a report entitled Occupational Licensing and Public Policy, which raises these issues. It was the only up-to-date and comprehensive document I was able to locate and it provided an excellent overview in itself of licensing practices in various occupations and their dubious quality.

Third, the civil rights movement has continued to make inroads against discrimination, specifically here concerned with discriminatory practices in hiring and occupations access. Equal Employment Opportunity Commission Guidelines, 1970, focuses attention on test validation in employment situations and there is reason to believe from various recent court decisions (such as Griggs vs. Duke Power Company, 401, U.S. 424, 1971) that the federal guidelines could be applied to licensing situations. The guidelines require that evidence of a test's validity:

> ". . .should consist of empirical data demonstrating that the test is predictive of or significantly correlated with important elements of work behavior which compromise or are relevent to the jobs for which candidates are being evaluated. Empirical evidence in support of a test's validity must be based on studies emphasizing generally accepted procedures, such as those described in Standards for Educational and Psychological Tests and Manuals, published by the American Psychological Association. However, evidence for content or construct validity should be accompanied by sufficient information from job analysis to demonstrate the relevance of the content or contruct."

The November "APA Monitor" clipping I included describes some recent extensions of the guidelines to local and state governments. (Our next speaker, Thomas Goolsby, Jr., will bring us up to date and discuss the legal questions further).

Fourth, are challenges to access being made to many professions to obtain status through alternatives to the traditional curriculum/school based training routes. This becomes a question of who qualifies to take a licensure exam? Are such exams really to protect the public or a limit access by those who have already made it? If exams are not proven valid in terms of job needs and as they are in most cases controlled by the professions themselves, then this is a meaningful issue for those who seek entry through alternative routes. For example, cases as reported of returning army medics who sought to take the RN licensure exam were denied on the grounds of not having graduated from nursing school.

It can be said that licensing is going through a period of questioning. For a number of reasons including questions of federal legality, licensing agencies are apt to soon be challenged to prove their tests are valid predictions of job performance significantly measuring job-related skills. It seems unlikely that any less will be acceptable.

Availability of Information

Despite a growing concern for licensure and validation in particular, there is a surprising lack of information and research on the topic. This is especially true in attempting to relate licensure to job performance. The information I was able to locate on licensing and related performance testing was scanty, often in progress, and done in subject matter areas rather than considered collectively as a methodological problem. In many cases, material was not available through generally accessible professional media and in some cases, professions considered such information confidential.

(This lack of information encouraged me to include with this paper the complete bibliography from my NLN study, hoping to save you the considerable trouble I went through in collecting sources.)

Maslow (1971), (one of our discussants) who was at the time with the Civil Service Commission Research Center, advised the Council on Occupation licensing:

> "I am convinced that we need to sharpen our ability to develop and demonstrate the rational relationship between the job requirements and the measurement system used to certify or qualify people for an occupation. A number of techniques are available to improve the process of job analysis to get a much more exact fix on the critical requirements for the work to be done. I would urge, therefore, that especially in examinations for occupational knowledge and proficiency, you insist, at the very least, on a clearcut showing of how one proceeds from the decision as to the skills and abilities required for effective performance to the decisions that certain tests or other measures will insure that the applicant can adequately perform in that occupation."

Validation Studies: The Problem

How have licensure validation studies been done? How should they be done? What do the studies available tell us? (This audience need not be reminded of the four generally accepted types of validity.)

Validation studies of licensure exams are rare. Seldom is the test development process that sophisticated or comprehensive. Many occupational groups call in teachers of their trade and/or practitioners at some point in the test development process. At worst, it is a rubber stamp operation. At best, it can approach a content validation methodology, but the quality of the process is limited by the adequacy of the universe specifications or how well the content from which the sample or test is drawn is defined and described. A second limiting factor has to do with how systematically the comments are requested, recorded and used. Such exercises are seldom reported except to say that they exist.

In my opinion, predictive criterion-related validation studies are the type most closely fitting the expressed purposes of licensure exams, that of assuring minimal competency on the job for the protection of the public. (The second "APA Monitor" clipping I have attached speaks to some professional disagreement on this matter). Concern is with a criterion not yet obtainable at the time of testing and one wishes to predict an individual's outcome prior to that situation occurring. They are 'selecting out' tests, as licensure was previously defined. Clearly, this suggests a research problem in itself, as those who fail are kept from practice and usually are not considered part of a validity study, as they are not practicing and available for observation in that job.

However, the major problem in predictive studies is finding appropriate job-related criteria. As Anastasi (1972) said:

> "Insofar as predictors are evaluated on the basis of their criterion measures, a validation study can be no better than the quality of its criterion data. Yet, in real-life situations, good criterion data are hard to come by."

Shimberg and others (1972) cite a similar, added logistical problem in regard to validation of licensure tests:

> "Individuals are licensed by a board, but once licensed they work for different employers--possibly in widely scattered locations. Any board that seeks to validate its tests by following up on the performance of each licensee faces a formidable task."

I think it can fairly be said that validation studies of the predictive type demanding job-related criteria are difficult to develop, time-consuming, impractical and expensive to perform. Psychometric methodology offers little guidance for such validation studies. The area of licensure in particular lacks the "classic" studies familiar to those schooled in psychological testing. Once this is comprehended, the fact that such validations are rare, almost non-existant, is less surprising but nevertheless disconcerting.

An Example of Validity Evidence: RN Licensure

Nursing was selected as the occupational example with which I am most familiar; and because on many of the findings in the Occupational Licensing and Public Policy report referred to above, the exam for RN licensure would rate high in comparison with other licensure exams reported upon. It is developed according to accepted test procedures, given under carefully controlled conditions, scored objectively and serves all states. To illustrate by comparison, some occupations build tests upon available text book questions (barbering) or make choices from a local file of essay questions (merchant marines). Most local or state exams have no reciprocity arrangements.

The RN licensure exam has never been directly validated, though rather typical content checks by nurse educators are routinely done. However, two types of studies are available which used the licensure exam as the criterion data, those that use the exam scores as a criterion variable in validity studies of other nursing tests, and studies which attempted to predict directly RN licensure scores. It is easier to use success on the licensure exam than to determine what constitutes success on the job or build an instrument to cover a multitude of job situations. For this reasons, the NLN uses the licensure exam to validate the predictive use of their pre-nursing exam. A high degree of relationship is found between the two. The RN licensure exam also correlates highly with the NLN achievement tests. However, a number of smaller studies, less definite but fairly consistant found that through theory grades were good predictors of licensure scores, clinical course grades were not; and correlations between theory and clinical course grades were lower than expected.

One can say with some confidence then, that the RN licensure exams are highly related to academic achievement but are such achievement measures necessarily related to the minimum competency required for the practice of beginning nursing? Obviously there is a necessary cognitive knowledge component to any

job but is it sufficient? "Is it possible, for instance," as one researcher asks (Taylor and other, 1966), "that students who do better in clinical practice courses than in more traditional academic classes will be more successful in actual work situations?" If in this or other fields licensure exams are more related to academic success than job performance, such findings will not only require changes in the licensure exams but more far reaching questioning of the curriculum and of the underlying occupational structure.

Testing Research and Job Performance

What does testing research suggest concerning the predictive validity of paper and pencil tests which are known to be highly related to success in school, school curriculum or academic grades? World War II Naval research is commonly credited as the point at which it became recognized that paper and pencil tests, though highly correlated with final course grades, were not effecient predictors of job performance.

> "Although it had been assumed that written tests sufficed to indicate what a man had learned in a service school, the evidence showed that performance tests and improved shop grades were not closely correlated with written test grades. During tryout in Gunners' Mates School, performance tests correlated from .14 to .35 with written tests and only slightly higher with final grades which were based largely on written tests." (Stuit, 1947)

These same written tests were also found to correlate well with reading tests (Gulliksen, 1950). Efforts were made following these findings to introduce more practical work and performance testing to the training.

This lack of relation between achievement as measured by traditional paper-and-pencil tests and performance measures, which appears in studies as diverse as education (Quirk and others, 1972) and engineering (Hemphill, 1963) suggests the great importance of test validation for licensure and certification. Although much lip service is given to the concept, it is seldom performed in an acceptable manner.

Ryans and Fredericksen (1951) sum up this point from a measurement perspective:

> "From the standpoint of validity one of the most serious errors committed in the field of human measurement has been that which assumes the high correlation of knowledge of facts and principles on the one hand and performance on the other. Nevertheless, examinations for admission to the bar, for medical practice, for teaching. . .are predominantly verbal tests of fact and principle in the respective fields."

If training and knowledge variables are not necessarily sufficient to define job proficiency, where does one look?

Performance Testing: Examples and Development

If one accepts Fitzpatrick's and Morrison's (1971) definition of performance testing as a test which is relatively realistic, then it is logical to look here for the answer to our questions of (1) how to validate licensure exams more effectively and (2) how to revise licensure tests if necessary.

The most interesting and well documented use I found concerning performance measures in predictive validation research was in the area of employee selection and promotion. Besides the monetary incentive for making a correct decision, an employers' situation has numerous advantages over licensure boards, such as control over subjects, the limited range of jobs and job descriptive information, and the possibility of gradually implementing a testing program, allowing research time to study predictions without actually implementing them.

Assessment centers are a performance-based type of employment or promotion screening device. The technique was originally devised to select secret service agents during World War II and applied in industrial situations by AT&T in the fifties. The procedures (Byham, 1970):

> ". . .simulates 'live' the basic situations with which a manager would be faced if he were moved up and develops information about how well he will cope at the higher level before the decision to promote him is actually made."

The assessors at the centers are trained observers, the exercises are standardized, and the conditions are constant and relatively realistic. This allows more valid comparative judgments to be made than in the 'real world'.

Two kinds of validity studies have been done. In an experimental setting, reports of the assessment are not released to management; thus no decisions are made on the basis of the assessment. The predictions are then compared with actual performance by some rating and/or observation technique, and other indicators of job success. If reports are released, which is more common but less conducive to sound validation, studies are then based on comparing those promoted before assessment center results were available to those promoted with this information, or by simply comparing progress of candidates promoted using assessment center reports and subsequent performance. According to Byham, all validation methods have tentatively pointed to the same conclusion:

> "The assessment center technique has shown itself a better indicator of future success than any other tool management has yet devised."

For a more descriptive example of how one such center works and the validation process was given by Bray and Campbell of AT&T (1968). Though the assessment center concept could be used as a validation tool, as an ongoing technique for licensure examinations it is obviously unrealistic.

To illustrate a more practical approach to introducing reality into the testing situation the medical profession has developed two types of programmed testing of clinical competence to simulate performance on objectively scored paper and pencil tests. The National Board of Medical Examiners first introduced the concept (Hubbard, 1964) and now uses programmed testing for the medical licensing exam Part III on clinical competence, which previously was a practical bedside type of oral exam. This is a linear model, while certification speciality exams use a branching model (McGuire and Babbott, 1967). In both, the examinee is confronted by a realistic clinical situation and procedes through a series of

decision choices, each step accompanied by an increment of information upon which the next depends, similar to programmed teaching. In the branching model the difference is that decision choices change, based upon previous choices allowing more than one route to a solution.

Neither variation has been validated in relation to predicting on the job performance but some work is in progress. The Part III or clinical competence exam is said to derive its validity from, among other things, measuring something different from Parts I and II, related to medical school course work. Cronbach (1970) having reviewed the validity evidence on Part III notes: "Follow up studies are needed to be sure that the test measures a skill of medical practice and not just ingenuity in test taking."

(Other examples of performance tests can be found in The Handbook of Performance Testing by Boyd and Shimberg (1971), although most are of a mechanical technical, variety).

Similarly to the problems confronted by those attempting predictive validation of licensure tests, performance test development logically begins with a study of specific skills and abilities involved in the activities the test is designed to measure or predict. The next step is the choice of representative tasks which strongly influences the validity of the performance test(s). Other difficulties with performance testing come from a lack of applied methodology in that performance tests are by nature criterion-referenced, and procedures for estimating reliability and validity are meager.*

Most literature on performance tests discuss them as a new form of assessment

---

* Licensure and certification exams have been discussed as types of minimal competency exams and like performance measures would normally be considered criterion referenced. The examinee is theoretically tested in terms of an absolute criteria, and comparisons among test takers is not a licensing purpose. However, most licensing tests are developed on norm-referenced models and the purposes. (I hope Robert Frary will bring this point into his discussion of methodology).

used to increase the realism of the test. The primary interest in performance tests expressed here is less commonly discussed, that of providing the criteria for predictive validation. The only description of such a research use I located was a theorectical discussion on "Providing a Criterion Measure" by Ryan and Frederiksen, 22 years ago (1951):

> "When the behavior involved in a situation is broad enough and representative enough of the situation as a whole, the performance is itself the criterion behavior for that situation. Consequently, performance test data, particularly when they refer to work samples, provide a more satisfactory measure of criterion behavior than is usually available. Because performance tests serve as a measure of the criterion, they may be of use in several important ways.
>
> Performance test data may provide, first of all, a criterion for research. Information yielded by performance tests makes possible the validation of other measures which, although of a more indirect nature, may be more economical in administration. In many situations it is difficult and expensive to administer performance tests to large numbers of examinees. Such situations demand the construction of psychometric instruments that will yield measurements related to criterion and will also be practicable. In the construction of aptitude tests for various skills and operations, performance tests may provide the criterion against which the available second order test can be judged."

As methodologists, the validation of licensing and certification exams presents real and immediate challenges. Here are practical problems, based on real and current concerns. If each occupation continues struggling on its own, without serious attempts from a group as we have here today, to provide an integrated conceptual and methodological framework, solutions will remain a long way off.

"APA Monitor" November, 1973

Testing and Equal Employment Opportunity

With the establishment of the Equal Employment Opportunity Commission (EEOC), and later the Office of Federal Contract Compliance (OFCC), there were powerful forces examining some of the discriminatory employment practices in both the public and private sectors. Aggrieved groups began to marshall the law in order to overcome the past effects of employment practices. Tests and test usage became the key issues in the development of these cases. Suddenly, terms which had been sacrosanct and strictly within the domain of psychology were being defined by opinions of judges in court cases. At the beginning of a case, a judge might have believed that the validity of a test depended on the presence of a stamp, but in the end, by the opinion he handed down, he was defining construct, content, or criterion-related validity based on the construction of the test.

Governmental guidelines were drawn up by both the OFCC and the EEOC to apply to all instances of test usage in employment. Although these Guidelines cited the APA *Standards*, they clearly stood on their own merits. In 1972, the Civil Rights Act was amended to give regulatory powers to the EEOC. This amendmest also established the Equal Employment Opportunity Coordinating Coencil (EEOCC) which is empowered with "the responsibility for developing anl implementing agreements policies and practices designed to maximize effort, promote efficiency, and eliminate conflict, competition, duplication and inconsistency among the operations, functions and jurisdictions of the various departments, agencies and branches of the Federal government responsible for the implementation and enforcement of equal employment opportunity legislation, orders, and policies." Development of the Uniform Guidelines on Employee Selection Procedures" is sthe first significant cooperative effort of the Council composed of representatives from member agencies—Civil Rights Commission, Civil Service Commission, Department of Justice, EEOC and the Department of Labor. The 1972 amendment to the Civil Rights Act, in addition to creating the Council, puts local and state governments under the jurisdiction of the EEOC and spells out nondiscrimimnation requirements for the federal government. This means that as many as an additional 18 million employees are brought under protection of this Act and will be affected by implementation of these Guidelines. The Guidelines will also have a profound effect on test development and usage.

It is refreshing that the government is being pro-active in seeking the counsel of the psychological profession prior to the adoption of these Guidelines as part of official policy. I urge members of APA to review this important document and to make their views publicly known. You can receive a copy of the *Uniform Guidelines on Employee Selection Procedures* by writing directly to the Office of Scientific Affairs, APA, 1200 17th Street, N.W., Washington, D.C. 20036. You should then make your comments directly to one of the member agencies of the Equal Employment Opportunity Council.

The Council will hold an open meeting for psychologists to discuss the proposed *Uniform Guidelines in Employee Selection Procedures* in Washington, D.C. at the U.S. Civil Service Commission, Room 1304, 1900 E St. N.W. beginning at 9 AM on November 15, 1973. Anyone interested in making a public statement on the guidelines should contact Mr. David Rose, Employment Section, Civil Rights Division, U.S. Department of Justice, 550 11th St. N.W., Room 1138, Washington, D.C., (202) 739-3831.

Leona Tyler
APA President

"APA Monitor" ?, 1973

Civil Service, EEOC spar over test validity

Two key federal agencies expressed diametrically opposing views on the validation of employment tests during an APA-sponsored open hearing on the revised *Standards for Development and Use of Educational and Psychological Tests* held in Washington last March.

While praising the *Standards* in general, Dr. John S. Howland, director of the U.S. Civil Service Commission's Personnel Research and Development Center took exception with what he interpreted as an implied endorsement of criterion-related validity as "the preferred model."

"Very real constrictions on the practical usefulness of classic criterion-related validity make it less and less attractive in employment situations," said Howland. ". . . Construct validity deserves equal time and attention and in the long run may be the most useful strategy for developing generality for tests."

"Criterion-related validity studies are usually not appropriate in our employment situations were persons are hired from the top of a list down in descending score order. We must rely on a system in which relevant job knowledges, skills, and abilities, identified through careful job analysis, are used to identify content and construct domains appropriate for assessing job applicants."

Dr. William H Enneis, chief of Research Studies for the Equal Employment Opportunity Commission said a "major concern" of his agency "has been the shift away from clearly appropriate criterion-related validity investigations and the consequent attempt to justify employee selection procedures under concepts that have no definite psychological meaning or empirical legitimacy."

Enneis called the *Standards* treatment of criterion-related validity "excellent," and suggested that if test developers and users heeded it they would "in most situations, satisfy the requirements of the EEOC *Guidelines*."

"Construct validity," he added, "from an employment viewpoint, is extremely difficult to accept because claims have been made for it without one shred of evidence that the constructs purportedly measured by the test are actually the same as those allegedly required on the job."

Issues of validity will again be high on the agenda when the Joint Committee charged with revising the *Standards* meets in mid-May to consider the suggestions made by Howland, Enneis and other concerned parties who appeared at the Washington hearings and similar meetings in San Francisco and New Orleans.

The final document is due to be published this summer.

At the February New Orleans meeting sponsored by the American Educational Research Association and the National Council on Measurement in Education—the two organizations collaborating with APA in the revision effort—some educators recommended deletion of Section L of the *Standards* which deals with program evaluation.

They argued that program evaluation involves more than just testing and could not be treated properly within the limited time available before publication of the *Standards*. Others advocated a separate document on program evaluation.

B I B L I O G R A P H Y

VALIDATION OF THE STATE BOARD TEST POOL
EXAMINATION FOR RN LICENSURE:
Working Paper prepared for the
National League for Nursing
in conjunction with the
SBTPE Research Steering Committee
for the RN Examination

By Kathryn A. Hecht, Ed.D.
February, 1973

National League for Nursing
10 Columbus Circle
New York, New York

* Those entries so marked are cited in this paper.

BIBLIOGRAPHY

* Anastasi, A. "Technical Critique." In An Investigation of Sources of Bias in the Prediction of Job Performance: A Six Year Study. Proceedings of Invitational Conference. Princeton: Educational Testing Service, 1972.

* Andrews, E. "Certification." In Houston, W. R. and Howsan (Eds.), Competency-Based Teacher Education. Chicago: Science Research Associates, 1972.

Bailey, J. T., McDonald, F. J., and Klaus, K. E. An Experiment in Nursing Curriculums at a University. San Francisco: San Francisco Medical Center, University of California, 1971.

Baldwin, J. P., Mowbray, J. K., and Raymond, T. G. "Factors Influencing Performance on State Board Test Pool Examinations." Nursing Research, March-April, 1968, Vol. 17, pp. 170-2.

Bircher, A. U. "Nursing Licensure and Nursing Turf." Supervisor Nurse, May, 1972, Vol. 3:5, pp. 60-71.

* Boyd, J. L., Jr., and Shimberg, B. Handbook on Performance Testing. Princeton: Educational Testing Service, 1971.

Brandt, E. M., Hastie, B., and Schumann, D. "Predicting Success on State Board Examinations." Nursing Research, Winter 1966, Vol. 15:1, pp. 62-8.

Brandt, J. "Licensure, Pros, Cons and Proposals." Chart, May 1972, pp. 74-7.

* Bray, D. W., and Campbell, R. J. "Selection of Salesmen by Means of an Assessment Center." Journal of Applied Psychology, 1968, Vol. 52:1, pp. 36-41.

* Byham, W. "Assessment Centers for Spotting Future Managers." Harvard Business Review, July-August 1970, Vol. 48:4, pp. 150-67.

* Cronbach, L. J. Essentials of Psychological Testing. (Third Edition) New York: Harper and Row, 1970.

Cronbach, L. J. "Test Validation." In Thorndike, R. L. (Ed.), Educational Measurement. Washington, D.C.: American Council on Education, 1971.

Educational Testing Service. Bulletin of Information: General Automobile Mechanic. Princeton: 1972.

Educational Testing Service. An Investigation of Sources of Bias in the Prediction of Job Performance: A Six Year Study. Proceedings of Invitational Conference: Educational Testing Service, 1972.

Fine, J. L., Malfette, J. L., and Shoben, E. J. The Development of Criterion for Driver Behavior. New York: Columbia University Press, 1965.

* Fitzpatrick, R., and Morrison, E. J. "Performance and Product Evaluation." In Thorndike, R. L. (Ed.), Educational Measurement (2nd edition). Washington, D.C.: American Council on Education, 1971.

Fivars, G., and Gosnell, D. Nursing Evaluation: The Problem and the Process: The Critical Incident Technique. New York: Macmillan Co., 1964.

Flanagan, J. C. "The Critical Incident Technique." Psychological Bulletin, 1954, Vol. 51, pp. 327-58.

Gorham, W. A. "Methods for Measuring Staff Nursing Performance." Nursing Research, Winter 1963, Vol. 12:1, pp. 4-11.

Gorham, W. A. "Staff Nursing Behaviors Contributing to Patient Care and Improvement." Nursing Research, Spring 1962, Vol. 2:2, pp. 68-77.

* Gulliksen, H. "Intrinsic Validity." American Psychologist, 1950, Vol. 5, pp. 511-17.

* Hemphill, J. K. A Summary of the Engineering Study. Princeton: Educational Testing Service, 1963.

Highland, R. W. A Guide for Use in Performance Testing in Air Force Technical Schools. Colorado: Lowry Air Force Base, 1955. (see Cronbach, 1971)

* Hoffmann, B. The Tyranny of Testing. New York: Crowell-Collier, 1962.

Hubbard, J. P. Measuring Medical Education. Philadelphia: Lea and Febiger, 1971.

* Hubbard, J. P. "Programmed Testing in the Examinations of the National Board of Medical Examiners." In 1963 Invitational Conference on Testing Problems. Princeton: Educational Testing Service, 1964.

* Jensen, O. "Some Thoughts on Minimum Competency Interpretations of Performance on (M-C Item) Achievement Tests." (mimeo), Educational Testing Service, March 1972.

Journal of the American Medical Association (JAMA), "Medical Licensure Statistics--1970." JAMA, June 14, 1971, Vol. 216, p. 11.

Klaus, D. J., Gosnell, D. E., Chowla, M. C., and Reilly, P. C. Controlling Experience to Improve Nursing Proficiency: Vol. III. Determining Proficient Performance: Pittsburgh: American Institutes for Research, 1968.

Klaus, D. J., Gosnell, D. E., Jacobs, A. M., Reilly, P. C., and Taylor, J. A. Controlling Experience to Improve Nursing Proficiency: Vol. I. Background and Study Plan. Pittsburgh: American Institutes for Research, 1966.

Ledbetter, P. J. "An Analysis of the Performance of Graduates of a Selected Baccalaureate Program in Nursing with Regard to Selected Standardized Examinations." Dissertation Abstracts. University Microfilms: Ann Arbor. Vol. 29, p. 3381-A, 1969. (see Litwack, 1972)

Lewy, A., and McGuire, C. "A Study of Alternative Approaches in Estimating the Reliability of Unconventional Tests." (presented at A.E.R.A., 1966) Center for the Study of Medical Education. Chicago: University of Illinois College of Medicine, 1966.

Litwack, L., Sakata, R., and Wykle, M. Counseling Evaluation and Student Development. Philadelphia: W. B. Saunders Co., 1972.

Magnusson, D. Test Theory. Reading, Massachusetts: Addison-Wesley Publishing Co., 1966.

*Maslow, A. P. "Licensing Tests-Occupational Bridge or Barrier?" Proceedings Third Annual Conference. Council on Occupation Licensing, July 1971. (see Shimberg, 1972)

*McDonald, F. J. "Evaluation of Teaching Behavior." In Houston, W. R., and Howsan (Eds.), Competency-Based Teacher Education, Chicago: Science Research Assoc., Inc., 1972.

McGuire, C. H. "An Evaluation Model for Professional Education--Medical Education." 1967 Invitational Conference on Testing Problems. Princeton: Educational Testing Service, 1968.

McGuire, C. H. Clinical Simulations. Appleton Century Crofts, 1972.

*McGuire, C. H., and Babbott, D. "Simulation Techniques in the Measurement of Problem-Solving Skills." Journal of Educational Measurement, Spring 1967, Vol. 4:1.

Mueller, E. J., and Lyman, H. B. "The Prediction of Scores on the State Board Test Pool Examination." Nursing Research. May-June 1969, Vol. 18:3, pp. 263-7.

National League for Nursing. "Analysis of State Board Test Scores 1957." Nursing Outlook. 1959, Vol. 7:5.

National League for Nursing. (1963) In Katzell, M. "Comments on Moll's Report." NLN 1964

National League for Nursing. "Safety and Effectiveness of Practice." Nursing Outlook ("Let's Examine" series), October 1962, Vol. 10.

National League for Nursing. "Standardization, Scoring and Use of Scores on the SBTPE." SB162-28, pp. 7-11.

*National League for Nursing. "Where Those Tests Come From and What They're For." Nursing Outlook ("Let's Examine" series), April 1961, Vol. 9.

Nealy, S. M., and Owen, T. W. "A Multitrait-Multimethod Analysis of Predictors and Criteria of Nursing Performance." Organizational Behavior and Human Performance, 1970, Vol. 5, pp. 348-65.

Palef, S., and Stewart, C. P. "Development of a Rating Scale." Studies in Personnel Psychology, October 1971, Vol. 3:2, pp. 7-20.

Popham, J. W. "Performance Tests of Teaching Proficiency." American Educational Research Journal, January 1971, Vol. 8:1, pp. 105-18.

Popham, W. J., and Husek, T. R. "Implications of Criterion-Referenced Measurement." Journal of Education Measurement, Spring 1969, Vol. 6:1, pp. 1-9.

Quirk, T. J. "Performance Tests for the Beginning Teachers: Why All the Fuss?" (Invited address, N. J. Performance Evaluation Project 1971), Princeton: Educational Testing Service.

* Quirk, T. J., Witten, B. J., and Weinberg, S. F. A Critical Review of Research Related to the National Teacher Examinations, Princeton: Educational Testing Service, 1972.

Rosen, A., and Abraham, G. E. "Evaluation of a Procedure for Assessing the Performance of Staff Nurses." Nursing Research, February 1963, Vol. 14, pp. 78-82.

* Ryans, D. G., and Frederiksen, N. "Performance Tests of Educational Achievement." In Linquist, E. F. (Ed.), Educational Measurement, Washington, D.C.: American Council of Education, 1951.

Schmidt, S. "Special Background Presentation." A presentation at the April 1972 meeting of the SBTPE Research Steering Committee for the RN Examination. (NLN) July, 1972.

Schorr, T. M. "A Critical Balance." American Journal of Nursing, November 1971, Vol. 71:11, p. 2127.

Shields, M. R. "A Project for Curriculum Improvement." Nursing Research, October 1952, Vol. 1:2, pp. 4-13.

* Shimberg, B., Esser, B. F., and Kruger, D. H. Occupational Licensing and Public Policy, Princeton: Educational Testing Service, October 1972.

Slater, D. "The Slater Nursing Competencies Rating Scale." Detroit: College of Nursing: Wayne State University, 1967.

Smith, P. C., and Kendall, L. M. "Retranslation of Expectations: An Approach to the Construction of Unambiguous Anchors for Rating Scales." Journal of Applied Psychology, 1963, Vol. 47:2, pp. 149-55.

* Stuit, D. B. (Ed.) Personnel Research and Test Development in the U.S. Bureau of Naval Personnel. Project N-106 and the College Entrance Examination Board. Princeton: Princeton University Press, 1947. (see Boyd, 1972)

* Taylor, C. W., Nahm, H., Loy, L., Harms, M., Berthold, J., and Wolfer, J. Selection and Recruitment of Nurses and Nursing Students: A Review of Research Studies and Practices. Salt Lake City: University of Utah Press (rev.), 1966.

Taylor, C. W., Nahm, H., Quinn, M., Harms, M., Mulaik, J., Mulaik, S. Report of Measurement and Prediction of Nursing Performance. Part I. Factor Analysis of Nursing Students' Application Data, Entrance Test Scores, Achievement Test Scores and Grades in Nursing School. Salt Lake City: University of Utah, 1965.

Tiffin, J., and McCormick, E. J. Industrial Psychology (Fifth Edition). Englewood Cliffs: Prentice-Hall, 1965.

*U.S. Government. Guidelines for Employee Selection Procedures Title 29 Labor, Chapter XIV, Equal Employment Opportunity Commission Federal Register. August 1970, Vol. 149:35.

Wandelt, M. A., and Ager, J. Quality Patient Care Scale. Detroit: College of Nursing, Wayne State University, 1970.

Wrigley, M. H. "A Study of the Relationship Between the Scores of Practical Nurses on the Licensing Examination and Ratings on Their Performance on the Job." Dissertation Abstracts. University Microfilms: Ann Arbor. Vol. 30, p. 1763-A, 1969.